Theoretical Computer Science 9 (1979) 99–125

THE NUMBER OF REGISTERS REQUIRED FOR EVALUATING ARITHMETIC EXPRESSIONS*

P. FLAJOLET, J.C. RAOULT and J. VUILLEMIN
Iria–Laboria, 78150 Rocquencourt, France
Université de Paris-Sud, Batiment 490, 91405 Orsay, France

Communicated by M. Nivat
Received January 1978
Revised May 1978

Abstract. We study the number of registers required for evaluating arithmetic expressions. This parameter of binary trees appears in various computer science problems as well as in numerous natural sciences applications where it is known as the Strahler number.

We give several enumeration results describing the distribution of the number of registers for trees of size n. The average number of registers has the asymptotic expansion $\log_4 n + D(\log_4 n) + o(1)$; here, function D is periodic of period one, and its Fourier expansion can be explicitly determined in terms of Riemann's zeta function and Euler's gamma function.

1. Evaluation of arithmetic expressions and exploration of trees

We propose here to study a function from binary trees to integers. This function is called the *register* function, denoted Reg since it appears in register allocation problems. The same function also occurs in procedures for tree exploration and in natural sciences applications.

Definition. The register function is a function from binary trees to integers defined inductively by

$$\begin{cases} \mathrm{Reg}(\square) = 0 \\ \mathrm{Reg}\left(\begin{array}{c} \circ \\ L \quad R \end{array}\right) = \begin{array}{llll} \textbf{if} & \mathrm{Reg}(L) = \mathrm{Reg}(R) & \textbf{then} & \mathrm{Reg}(L) + 1 \\ \textbf{else} & & & \max(\mathrm{Reg}(L), \mathrm{Reg}(R)). \end{array} \end{cases}$$

For each tree, the register function is thus easily evaluated in a bottom-up way by assigning 0 to each external node and by progressing in the tree according to one of the two rules of the definition. The construction of such a labelling is examplified in Fig. 1.

* A preliminary version of this paper has appeared in the 18th Annual Symposium on Foundations of Computer Science, Providence, October 1977.

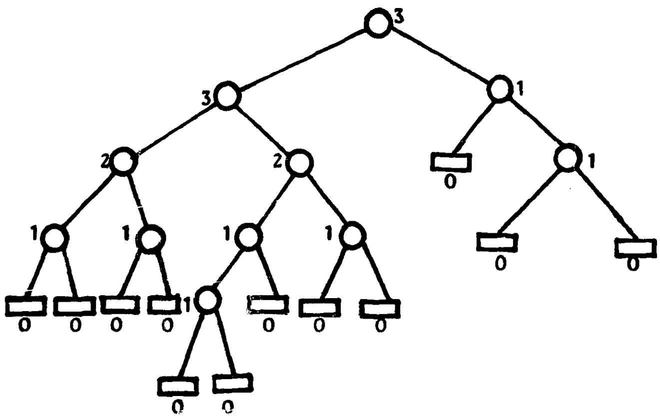

Fig. 1.

Notice that Reg(T) is also equal to the height of the largest complete subtree that can be embedded in T.

1.1. Evaluation of arithmetic expressions.

Consider an arithmetic expression, say $(x+y^z)\times t$.

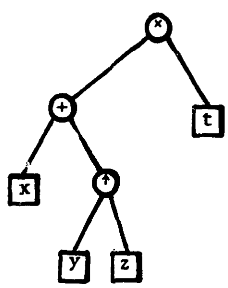

Fig. 2.

The traditional computer science representation of this expression is the binary tree shown in Fig. 2. In order to evaluate it, for given values of x, y, z, t, in a digital computer we have to use registers and produce code such as

$$\begin{aligned}
&\mathrm{R0} \leftarrow x\\
&\mathrm{R1} \leftarrow y\\
&\mathrm{R2} \leftarrow z\\
&\mathrm{R1} \leftarrow R1 \uparrow R2\\
&\mathrm{R0} \leftarrow \mathrm{R0}+\mathrm{R1}\\
&\mathrm{R1} \leftarrow t\\
&\mathrm{R0} \leftarrow \mathrm{R0}\times \mathrm{R1}
\end{aligned}$$

which leaves the desired result in register R0, using the two registers R1 and R2 for storing intermediate values. For this particular expression, it is also possible to use the code:

$$\begin{aligned}
&\mathrm{R0} \leftarrow y\\
&\mathrm{R1} \leftarrow z\\
&\mathrm{R0} \leftarrow \mathrm{R0} \uparrow \mathrm{R1}\\
&\mathrm{R1} \leftarrow x\\
&\mathrm{R0} \leftarrow \mathrm{R0} + \mathrm{R1}\\
&\mathrm{R1} \leftarrow t\\
&\mathrm{R0} \leftarrow \mathrm{R0} \times \mathrm{R1}
\end{aligned}$$

which is better than the previous one since it uses one register less. It has been shown as early as 1958 by Ershov that there exists an optimal strategy (w.r.t. the number of registers) for the computation of general expressions formed with binary operators; the optimal evaluation of expression E uses one register for storing the final result, and a certain number of extra registers, for keeping intermediate results. This number is exactly the register function of the tree underlying E, and it will also be denoted $\mathrm{Reg}(E)$.

If E is an expression involving n operators, i.e. its binary tree representation has n internal nodes ○ and $(n+1)$ leaves □, it is straightforward to show that $\mathrm{Reg}(E) \leq \lfloor \log_2 (n+1) \rfloor$.

Assuming all binary trees with n internal nodes to be equally probable, we are set to study the average number A_n of registers required for evaluating expressions of size n, i.e. involving n operators and $(n+1)$ operands.

For example, if $n=3$ we have Fig. 3.

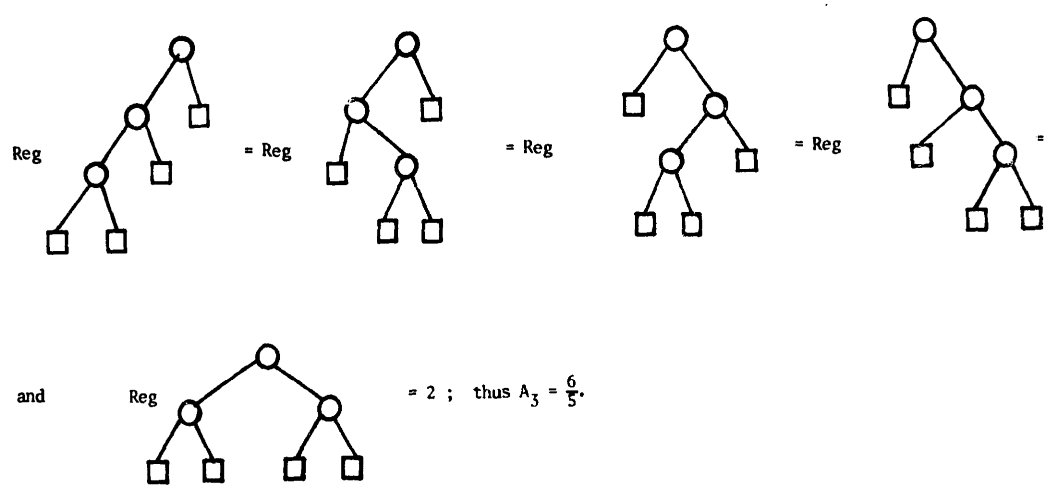

Fig. 3.

1.2. Traversing binary trees

This same quantity also arises in the analysis of an interesting algorithm for traversing (i.e. visiting each node of) binary trees.

One popular method for visiting binary trees has been named preorder in [11] and can be described by the following ALGOLish procedure.

```
proc PREORDER(T)
    {binary tree T}
    if T ≠ ∅: VISIT(ρ(T)); PREORDER(l(T)); PREORDER(r(T)) fi
cproc
```

In this notation, T is a binary tree, $\rho(\tau)$ represents its root, $l(T)$ and $r(T)$ its left and right subtrees; the procedure VISIT performs some action on each of the nodes of T. Such a recursive procedure can be implemented non-recursively with the help of a stack (see [6, 11, 12]).

```
proc PREORDER(T)
    {binary tree T, stack S}
    do do while T ≠ ∅:
            VISIT (ρ(τ)); S ⇐ r(T);
            T ← l(T)
       od
       while S ≠ ∅: T ⇐ S
    od
cproc
```

Here ($S \Leftarrow a$) pushes a on the stack S, ($a \Leftarrow S$) pops the top of the stack, and the control structure **do** $B1$ **while** $p : B2$ **od** is equivalent to ALGOL's $B1$; **while** p **do begin** $B2; B1$ **end**.

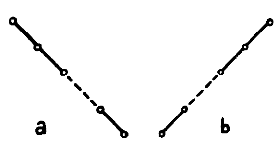

Fig. 4.

The parameter we are worried about in the analysis of this algorithm is the size of the stack S. For a tree of size n like Fig. 4a, this number is equal to one, while it is equal to n for Fig. 4b. Although it has been shown in [1] that the average size of the stack is only

$$\sqrt{\pi n} - \frac{1}{2} + O\left(\frac{\log n}{\sqrt{n}}\right),$$

assuming all binary trees of size n to be equally probable, there are applications for which such an extra storage requirement is prohibitive.

In order to cut down on the stack size, rather than stick to a fixed order in which to enumerate subtrees, left then right in our example, it is interesting to explore the "cheap" subtree first, and the "expensive" one later. Here, "cheap" refers to the stack size required to explore the corresponding subtree. Precisely, we define the EORDER (for economical order) traversing of binary trees by the following program.

proc EORDER(T)
  {**binary tree** T}
  if $T \neq \emptyset$: VISIT($\rho(T)$);
    if Reg($l(T)$) $\leqslant$ Reg($r(T)$): EORDER ($l(T)$); EORDER ($r(T)$)
    elsif Reg ($l(T)$) $\geqslant$ Reg($r(T)$): EORDER ($r(T)$); EORDER ($l(T)$)
    fi
  fi
cproc

The EORDER algorithm explores a tree T with a stack of size Reg (T) a quantity bounded above by $\log_2(|T|+1)$. It can be used in contexts where the information represented by the value of Reg can be dynamically maintained.

1.3. Natural sciences applications

The number of registers also plays an important role in understanding the common rules that apply to tree structures one can find in nature. It has been used to explain quantitative similarities between such fundamentally different kinds of ramification as branching in trees, watershed drainage networks, blood vessels and bronchial tubes [15]. For pulmonary arteries, bronchial networks and several species of trees, the diameter of a segment in the structure is proportional to the number of registers of the underlying binary tree rooted at this segment. In natural problems, this quantity is generally known as the Strahler number [22]. For trees such as the sequoia, the Douglas fir and the ponderosa pine, the Strahler number is proportional to the height of the tree. References to work in natural sciences pertaining to the Strahler number were kindly provided to us by E.M. Reingold and are included in the bibliography.

2. Explicit enumerations

The aim of this section is to derive exact results on the number $R_{p,n}$ of binary trees of size n whose evaluation requires p registers.

In Section 2.1, the numbers $R_{p,n}$ are introduced together with some related quantities. The inductive definition of the function Reg translates into a recurrence relation on the $R_{p,n}$.

Section 2.2 introduces the generating functions $R_p(z)$ associated with the sequences $(R_{p,n})_{n\geq 0}$. The recurrence on the $R_p(z)$ is solved first by means of a trigonometric change of variable. We then derive the expression of $R_p(z)$, which turns out to be a rational fraction involving the Tchebycheff polynomials. These computations also bring out to light unexpected connections with results on the height of trees obtained in [1].

In Section 2.3, we come back to the coefficients $R_{p,n}$; by decomposing the rational function $R_p(z)$ into simple elements and expanding, we get a trigonometric expression for $R_{p,n}$ as a sum of powers of cosines.

Algebraic manipulations on $R_{p,n}$ are then used to derive an equivalent expression as a sum of binomial coefficients.

2.1. Definitions and basic recurrences

We first need review some results pertaining to binary trees.

Let B_n denote the number of binary trees with n nodes. The value of B_n is well-known (cf. [11] for a bibliography). A binary tree with n nodes is defined by its root, its left subtree with n_1 nodes and its right subtree with n_2 nodes, hence the recurrence relation

$$B_n = \sum_{n_1+n_2+1=n} B_{n_1}B_{n_2} \qquad n \geq 1,$$

$$B_0 = 1.$$

Considering the generating function

$$B(z) = \sum_{n\geq 0} B_n z^n,$$

the recurrence relation yields the algebraic equation $B-1=zB^2$, whose only solution in the ring of power series is

$$B = \frac{1-\sqrt{1-4z}}{2z}.$$

This gives an explicit form $B = 1+z+2z^2+5z^3+14z^4+\cdots+B_nz^n+\cdots$, where $B_n = \binom{2n}{n}/n+1$ is the nth Catalan number.

We now consider the number $R_{p,n}$ of trees of size n whose evaluation requires exactly p registers. From section 1, we know that for fixed n $R_{p,n}=0$ provided $p > \lfloor \log_2(n+1) \rfloor$.

It is easy to evaluate $R_{1,n}$: any tree requiring one register is necessarily a chain, which is determined by $n-1$ left/right choices, so $R_{1,n} = 2^{n-1}$.

Since the only tree of size 2^p-1 requiring p registers is perfectly balanced of height $(p+1)$, we have $R_{p,2^p-1} = 1$.

We are also interested in the number $S_{p,n}$ of trees of size n whose evaluation requires at least p registers.

$$S_{p,n} = \sum_{j \geq p} R_{j,n}.$$

Finally we introduce the quantity M_n representing the total number of registers required for evaluating all expressions of size n.

$$M_n = \sum_{p \geq 0} pR_{p,n} = \sum_{p \geq 0} S_{p,n'}$$

and the average quantity we are looking for is M_n/B_n.

It follows from the inductive definition of Reg given in section 1 that

$$\mathrm{Reg}\left(\begin{smallmatrix} & \circ & \\ / & & \backslash \\ T_1 & & T_2 \end{smallmatrix}\right) = p$$

in either of the following three cases:

(i) $\mathrm{Reg}(T_1) = p-1$ and $\mathrm{Reg}(T_2) = p-1$.
(ii) $\mathrm{Reg}(T_1) = p$ and $\mathrm{Reg}(T_2) < p$,
(iii) $\mathrm{Reg}(T_1) < p$ and $\mathrm{Reg}(T_2) = p$.

The number of trees with a left subtree of size n_1, a right subtree of size n_2 requiring p registers is thus

$$R_{p,n} = \sum_{n_1+n_2+1=n} \left(R_{p-1,n_1} \cdot R_{p-1,n_2} + R_{p,n} \cdot \sum_{j<p} R_{j,n_2} + \sum_{j<p} R_{j,n_1} \cdot R_{p,n_2} \right)$$

$$= \sum_{n_1+n_2+1=n} \left(R_{p-1,n_1} \cdot R_{p-1,n_2} + 2R_{p,n_1} \cdot \sum_{j<p} R_{j,n_2} \right).$$

This relation gives $R_{p,n}$ from the set of $R_{j,k}$ preceding $R_{p,n}$ in the lexicographical ordering of N^2, thus allowing computation of all $R_{p,n}$'s

Table 1
$R_{p,n}$ for small values of p and n

$p \backslash n$	1	2	3	4	5	6	7	8	9	10
1	1	2	4	8	16	32	64	128	256	512
2			1	6	26	100	364	1288	4488	15 504
3							1	14	118	780

2.2. *Generating functions*

In this section we exhibit several equivalent forms for the generating functions associated to the sequences $\{R_{p,n}\}_{n \geq 0}$, $\{S_{p,n}\}_{n \geq 0}$ and $\{M_n\}_{n \geq 0}$.

We first define

$$R_p(z)=\sum_{n\geq 0} R_{p,n}\cdot z^n,$$

$$S_p(z)=\sum_{n\geq 0} S_{p,n}\cdot z^n=\sum_{j\geq p} R_j(z)$$

and

$$M(z)=\sum_{n\geq 0} M_n\cdot z^n=\sum_{p\geq 0} S_p(z).$$

The recurrence relation on $R_{p,n}$ readily translates into

$$R_p=zR_{p-1}^2+2zR_p\cdot\sum_{j<p} R_j \qquad \text{for } p\geq 1 \text{ and}$$

$$R_0=1.$$

In particular

$$R_1=\frac{z}{1-2z}=z+2z^2+4z^3+8z^4+16z^5+\cdots,$$

$$R_2=\frac{z^3}{1-6z+10z^2-4z^3}=z^3+6z^4+26z^5+100z^6+\cdots.$$

Solving the recurrence above allows to derive expressions for the generating functions R_p, S_p, M.

Proposition 1. *The generating functions R_p, S_p and M have the forms*

$$R_p(z)=2i\sin\phi\,\frac{t^{2^p}}{1-t^{2^{p+1}}}$$

$$S_p(z)=2i\sin\phi\,\frac{t^{2^p}}{1-t^{2^p}}$$

$$M(z)=2i\sin\phi\sum_{n>0} v_2(n)t^n$$

where ϕ is related to z by $e^{-i\phi}=B(z)-1$, and $t=e^{-i\phi}$. The function $v_2(n)$ is the dyadic valuation[1] *of n.*

Proof. In order to eliminate the summation sign in the recurrence giving R_p, we divide both sides by zR_p, thus

$$\frac{1}{z}=2\cdot\sum_{j<p} R_j+\frac{R_{p-1}^2}{R_p}.$$

[1] See definition below.

Subtracting this from the analogous identity obtained for $p+1$ yields

$$0 = 2R_p + \frac{R_p^2}{R_{p+1}} - \frac{R_{p-1}^2}{R_p}.$$

Dividing by R_p, we get an expression involving only the ratios $U_p = R_{p-1}/R_p$:

$$\begin{cases} U_{p+1} = U_p^2 - 2 \\ U_1 = \dfrac{R_0}{R_1} = \dfrac{1}{z} - 2. \end{cases}$$

This particular form of quadratic recurrence can be solved by a trigonometric change of variables. Setting $U_1 = 2\cos\phi$ and in general $U_p = 2\cos\phi_p$, we see that

$$\begin{aligned} U_{p+1} &= 2\cos\phi_{p+1} = 4\cos^2\phi_p - 2 \\ &= 2\cos 2\phi_p. \end{aligned}$$

It follows that $\phi_{p+1} = 2\phi_p = 2^p\phi$, which gives the explicit form

$$U_{p+1} = 2\cos 2^p\phi \qquad \text{for } p \geqslant 1,$$

$$U_1 = 2\cos\phi = \frac{1}{z} - 2.$$

A trigonometric expression of R_p follows from computing

$$U_p \times \cdots \times U_1 = \frac{R_{p-1}}{R_p} \times \cdots \times \frac{R_0}{R_1} = \frac{1}{R_p} = (2\cos 2^{p-1}\phi)(\cdots)(2\cos\phi).$$

This product collapses when we multiply by $\sin\phi$

$$\begin{aligned} \frac{\sin\phi}{R_p} &= (2\cos 2^{p-1}\phi)\cdots(2\cos 2\phi)(2\cos\phi\sin\phi) \\ &= (2\cos 2^{p-1}\phi)\cdots(2\cos 4\phi)(2\cos 2\phi\sin 2\phi) \\ &= \cdots = \sin 2^p\phi. \end{aligned}$$

We thus have a very simple expression for R_p:

$$R_p = \frac{\sin\phi}{\sin 2^p\phi} \tag{2}$$

where

$$\cos\phi = \frac{1}{2z} - 1, \quad \text{or } z = \frac{1}{4\cos^2\dfrac{\phi}{2}}.$$

This last equality only determines ϕ up to its sign; using $\sin^2\phi + \cos^2\phi = 1$ yields

$$\sin^2\phi = \frac{1}{z} - \frac{1}{4z^2},$$

and we can choose either sign for sin ϕ, which determines ϕ completely. We can take for instance

$$\frac{\sqrt{1-4z}}{2z} = \mathrm{i} \sin \phi,$$

or equivalently

$$B = \frac{1}{2z} - \frac{\sqrt{1-4z}}{2z} = 1 + \cos \phi - \mathrm{i} \sin \phi, \quad \text{i.e.}$$

$$B - 1 = zB^2 = \mathrm{e}^{-\mathrm{i}\phi}.$$

We can now rewrite the expression for R_p as

$$R_p = 2\mathrm{i} \sin \phi \frac{t^r}{1-t^{2r}} \tag{2'}$$

where $t = \mathrm{e}^{-\mathrm{i}\phi}$, $r = 2^p$.

We get similar expressions for S_p and M, by summations:

$$\sum_{\substack{j \geq p \\ r = 2^j}} \frac{t^r}{1-t^{2r}} = \sum_{\substack{j \geq p \\ k \geq 0}} t^{2^j(2k+1)} = \sum_{\substack{m,k \geq 0 \\ u = t^{2^p}}} u^{2^m(2k+1)}.$$

Since $(m, k) \to 2^m(2k+1)$ is a bijective mapping $\mathbf{N}^2 \to \mathbf{N}^+$, this last expression equals $u/(1-u)$, thus

$$S_p = 2\mathrm{i} \sin \phi \frac{t^r}{1-t^r}, \tag{3}$$

where $t = \mathrm{e}^{-\mathrm{i}\phi}$, $r = 2^p$.

Elementary trigonometric manipulations yield the equivalent form:

$$S_p = -\mathrm{i} \sin \phi + \sin \phi \operatorname{cotg} 2^{p-1}\phi. \tag{3'}$$

Finally, the expression for $M = \sum_{p \geq 1} S_p$ follows from summing

$$\sum_{\substack{p \geq 1 \\ r = 2^p}} \frac{t^r}{1-t^r} = \sum_{p,k \geq 1} t^{k2^p}.$$

The application $(p, k) \to k2^p$ from $\mathbf{N} \times \mathbf{N}^+ \to \mathbf{N}^+$ is surjective; the cardinality of the inverse image of n is $v_2(n)$, the unique integer such that $n = 2^{v_2(n)} \cdot (2k+1)$ for $k \geq 0$, $n \geq 0$. The quantity $v_2(n)$ is known as the *dyadic valuation* of n; the numbers $v_2(n)$ can be computed by the rules:

$$v_2(2i+1) = 0 \qquad \text{for } i \geq 0 \quad \text{and}$$

$$v_2(2i) = 1 + v_2(i) \qquad \text{for } i > 0.$$

Table 2
The dyadic valuation $v_2(n)$

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$v_2(n)$	0	1	0	2	0	1	0	3	0	1	$\cdots$

The generating function M then takes the form:

$$M = 2\mathrm{i}\sin\phi \sum_{n>0} v_2(n)\,\mathrm{e}^{-\mathrm{i}n\phi}. \tag{4}$$

This completes the proof of Proposition 1.

Proposition 1 provides an expression of $R_p(z)$ in terms of powers of the algebraic function $B(z)$. Eliminating ϕ, we get

$$R_p = -\frac{1}{z}\sqrt{1-4z}\{(zB^2)^{2^p} - (zB^2)^{-2^p}\}^{-1}.$$

On the other hand simpler expressions may exist since we know that both $R_p(z)$ and $B(z) - S_p(z)$ are rational functions in z. Indeed we can show the following.

Proposition 2. *The generating functions R_p and S_p have the expressions*:

$$R_p(z) = \frac{z^{2^p-1}}{\bar{U}_{2^{p+1}}(z)}$$

$$S_p(z) = B(z) - \frac{\bar{U}_{2^{p-1}}(z)}{\bar{U}_{2^p}(z)},$$

where $\{\bar{T}_j\}_{j\geq 0}$ and $\{\bar{U}_j\}_{j\geq 0}$ are modified Tchebycheff polynomials.

Proof. It is known since Moivre that $\cos s\theta$ and $(\sin s\theta)/(\sin\theta)$ are polynomials in $\cos\theta$. Indeed

$$(\cos\theta + i\sin\theta)^s = \cos s\theta + \sin s\theta$$
$$= T_s(\cos\theta) + i\sin\theta U_s(\cos\theta).$$

in which T_s and U_s are the Tchebycheff polynomials with $\deg T_s = \deg U_{s+1} = s$.

Thus equations (2) and (3) yield rational expressions

$$R_p = \frac{1}{U_{2^p}(x)} \quad \text{with } x = \frac{1}{2z} - 1$$

$$S_p = B - \frac{1}{2z} + \frac{T_{2^{p-1}}(x)}{U_{2^p}(x)}$$

Simpler expressions are now derived; first express R_p and S_p in terms of $\frac{1}{2}\phi$:

$$R_p = \frac{\sin \frac{1}{2}\phi}{\sin 2^p\phi}\, 2\cos\tfrac{1}{2}\phi = \frac{2}{\sqrt{z}\cdot U_{2^{p+1}}\left(\dfrac{1}{2\sqrt{z}}\right)}$$

$$S_p = -\tfrac{1}{2}\sin\phi + 2\cos\tfrac{1}{2}\phi\cdot\frac{\sin\frac{1}{2}\phi}{\sin 2^{p-1}\phi}\cdot\cos 2^{p-1}\phi$$

$$= -\frac{1}{2z} + B(z) + \frac{1}{\sqrt{z}}\cdot\frac{T_{2^p}\left(\dfrac{1}{2\sqrt{z}}\right)}{U_{2^p}\left(\dfrac{1}{2\sqrt{z}}\right)}.$$

Classically, the T_i, U_i satisfy

$$T_{i+2}(x) + T_i(x) = 2xT_{i+1}(x); \qquad T_0 = 1; \qquad T_1 = x$$

$$U_{i+2}(x) + U_i(x) = 2xU_{i+1}(x); \qquad U_0 = 0; \qquad U_1 = 1$$

which are mere translations of trigonometric identities.
Now we can introduce the modified Tchebycheff polynomials

$$\bar{T}_i(z) = z^{i/2}T_i\left(\frac{1}{2\sqrt{z}}\right) \quad \text{and} \quad \bar{U}_i(z) = z^{(i-1)/2}U_i\left(\frac{1}{2\sqrt{z}}\right).$$

They satisfy the recurrences

$$\bar{T}_{i+2}(z) + z\bar{T}_i(z) = \bar{T}_{i+1}(z); \qquad \bar{T}_0 = 1; \qquad \bar{T}_1 = \tfrac{1}{2}$$

$$\bar{U}_{i+2}(z) + z\bar{U}_i(z) = \bar{U}_{i+1}(z); \qquad \bar{U}_0 = 0; \qquad \bar{U}_1 = 1.$$

In terms of $\bar{T}$ and $\bar{U}$, the preceding formulae take the form

$$R_p(z) = \frac{z^{2p-1}}{\bar{U}_{2^{p+1}}(z)},$$

and

$$S_p(z) = B(z) - \frac{1}{2z} + \frac{1}{z}\,\frac{\bar{T}_{2^p}(z)}{\bar{U}_{2^p}(z)} = B(z) - \frac{\bar{U}_{2^p-1}(z)}{\bar{U}_{2^p}(z)}.$$

These expressions together with the recurrence relations on the polynomials $\bar{T}_i$ and $\bar{U}_i$ permit a direct computation of R_p and S_p for each p.

At this point it should be noticed further that the coefficients of the polynomials $\bar{T}_i$ and $\bar{U}_i$ can be explicitly determined. Take for instance the $\bar{U}_i$ and consider the double generating function $Y(t, z) = \sum_{i\geq 0}\bar{U}_i(z)t^i$; it follows from the recurrence relation on the $\bar{U}_i$'s that $Y(t, z) = t/(1 - t + zt^2)$, hence $\bar{U}_{m+1}(z) = \sum_{k\geq 0}(-1)^k\binom{m-k}{k}z^k$.

In other words, the coefficients of the $\bar{U}_m$ are the upward diagonals of Pascal triangle taken with alternating signs.

These polynomials occur in several places in connection with tree enumerations. In particular they appear in the generating function of the number of trees of a given height [1, 13].

An explicit connection with these results is given by:

Proposition 3. *The number of binary trees with n internal nodes requiring less than p registers is equal to the number of general trees with* $(n+1)$ *nodes and height less than* 2^p.

Proof. De Bruijn, Knuth and Rice have shown that the number $A_{h,n}$ of (general) trees of height h and size n admits the generating function

$$A_h(z) = z\,\frac{\bar{U}_h}{\bar{U}_{h+1}}.$$

Elementary manipulations from Proposition 2 yield the result.

Notice also that we have the equality $B(z) = \sum_{p\geq 0} R_p(z)$, saying that each tree is assigned a uniquely defined register number, so that

$$B(z) = 1 + \sum_{p\geq 1} \frac{z^{2^p-1}}{\bar{U}_{2^{p+1}}(z)}.$$

Setting $z=-1$ in this equality, we see that $B(-1) = (\sqrt{5}-1)/2$ is the conjugate of the golden ratio; similarly $\bar{U}_{2^{p+1}}(-1)$ is the Fibonacci number of rank 2^{p+1}. Results of this section thus generalize a formula originally due to Lucas [14]:

Proposition 4. *The inverses of Fibonacci numbers whose indices are powers of two can be summed exactly*

$$\frac{\sqrt{5}-1}{2} = 1 - \frac{1}{F_4} - \frac{1}{F_8} - \frac{1}{F_{16}} \cdots .$$

2.3. Enumeration results

We now give expressions for the coefficients $R_{p,n}$, $S_{p,n}$ and M_n. Results from the last section make it possible to obtain partial fraction expansions of R_p and S_p, from which expansions as power series follow. This leads to expressions for the coefficients as sums of powers of cosines. Using elementary complex algebra, we derive equivalent expressions for $R_{p,n}$, $S_{p,n}$ and M_n in terms of binomial coefficients.

Proposition 5. *The quantities* $R_{p,n}$, $S_{p,n}$ *are given by*

$$R_{p,n} = \frac{4^n}{2^{p-1}} \sum_{1\leq k\leq 2^p} (-1)^{k+1} \sin^2 k\frac{\pi}{2^{p+1}} \cos^{2n} k\frac{\pi}{2^{p+1}}$$

and

$$B_n - S_{p,n} = \frac{4^{n+1}}{2^p} \sum_{1 \leq k < 2^{p+1}} \sin^2 k \frac{\pi}{2^p} \cos^{2n} k \frac{\pi}{2^p}.$$

Proof. In order to express R_p as a power series we decompose it as a sum of partial fractions:

$$R_p = \sum_k \frac{a_k}{x - s_k} \quad \text{where } x = \frac{1}{2z} - 1.$$

Here s_k is the kth root of U_{2^p-1}, that is $\sin 2^p \phi_k = 0 \Leftrightarrow \phi_k = k(\pi/2^p)$, and $s_k = \cos \phi_k$; it follows that

$$s_k = \cos k \frac{\pi}{2^p} \quad (k = 1, \ldots, 2^p - 1).$$

The coefficient a_k is given by

$$a_k = \frac{1}{U'_{2^p-1}(s_k)},$$

where U' denotes the derivative of $U(x)$. To obtain an explicit form, begin with

$$\frac{\mathrm{d}}{\mathrm{d}(\cos\phi)} \frac{\sin 2^p\phi}{\sin\phi} = \left(\frac{2^p \cos 2^p\phi}{\sin\phi} - \frac{\cos\phi \sin 2^p\phi}{\sin^2\phi} \right) \frac{1}{-\sin\phi}.$$

Substituting ϕ_k for ϕ, the second term vanishes, so that

$$\frac{\mathrm{d}}{\mathrm{d}x} U_{2^p-1}(s_k) = -2^p \frac{\cos 2^p\phi_k}{\sin^2\phi_k} = (-1)^{k+1} \frac{2^p}{\sin^2\phi_k}.$$

Hence we obtain for R_p the expression

$$R_p = \sum_{1 \leq k \leq 2^p-1} \frac{a_k}{1/2z - 1 - \cos\phi_k} = \sum_{1 \leq k \leq 2^p-1} \frac{2za_k}{1 - 4z\cos^2\phi_k/2}$$

with

$$a_k = \frac{1}{2^p} (-1)^{k+1} \sin^2 \phi_k.$$

Each term of the sum admits a power series expansion:

$$R_p = \sum_{1 \leq k \leq 2^p-1} 2za_k \sum_{n \geq 0} 4^n \cos^{2n} \frac{\phi_k}{2} z^n.$$

Identifying $R_{p,n}$ with the coefficient of z^n gives the value:

$$R_{p,n} = \frac{4^n}{2^{p-1}} \sum_{1 \leq k \leq 2^p-1} (-1)^{k+1} \sin^2 k \frac{\pi}{2^{p+1}} \cos^{2n} k \frac{\pi}{2^{p+1}}. \tag{6}$$

A very similar treatment can be applied to $B(z)-S_p(z)$ which ultimately leads to

$$S_{p,n}=B_n-\frac{4^{n+1}}{2^p}\sum_{1\le k\le 2^{p-1}-1}\sin^2 k\frac{\pi}{2^p}\cos^{2n}k\frac{\pi}{2^p}. \tag{7}$$

A trigonometrical form also exists for M_n which still involves the dyadic valuation v_2. These expressions are of interest since they give the asymptotic behaviours of the $R_{p,n}$ and $S_{p,n}$ for fixed p, as $n\to\infty$.

The next stage is to transform these sums of powers of cosines into sums of binomial coefficients. Similar equivalences are known since Ramus and they extend the method used to compute sums like $\binom{n}{0}+\binom{n}{2}+\binom{n}{4}+\cdots$.

Our start point is an equality readily checked by direct binomial expansion:

$$\sum_{0\le k<r}\omega^{-kt}(1+\omega^k)^m=r\sum_{j\in\mathbf{Z}}\binom{m}{t+jr},$$

where ω is a primitive rth root of unity. We use it to show

Theorem 1. *The quantities $R_{p,n}$, $S_{p,n}$ and M_n satisfy*

$$R_{p,n}=\sum_{\substack{j\in\mathbf{Z}\\ q=2^p(2j+1)}}\left[\binom{2n}{n-1+q}-\binom{2n}{n+q}\right],$$

$$S_{p,n}=-B_n+\sum_{\substack{j\in\mathbf{Z}\\ q=2^p\cdot j}}\left[\binom{2n}{n-1+q}-\binom{2n}{n+q}\right],\quad \textit{and}$$

$$M_n=\sum_{j\ne 0}v_2(j)\left[\binom{2n}{n-1+j}-\binom{2n}{n+j}\right].$$

Proof. One first translates the expression of $R_{p,n}$ given by Proposition 5. Let ω be a 2^{p+1} primitive root of unity; using obvious symmetries we get

$$\begin{aligned}R_{p,n}&=\frac{4^n}{2^{p-1}}\sum_{0\le k<2^{p+1}}(-1)^{k+1}\left(\frac{\omega^{k/2}-\omega^{-k/2}}{2i}\right)^2\left(\frac{\omega^{k/2}+\omega^{-k/2}}{2}\right)^{2n}\\&=\frac{1}{2^{p+1}}\sum_{0\le k<2^{p+1}}\omega^{k\cdot 2^p}(\omega^k-2+\omega^{-k})(\omega^k+1)^{2n}\omega^{-kn}\\&=\frac{1}{2^{p+1}}\sum_{0\le k<2^{p+1}}(\omega^k-2+\omega^{-k})\omega^{-k(n-2^p)}(1+\omega^k)^{2n}.\end{aligned}$$

Distributing $(\omega^k-2+\omega^{-k})$, the sum splits in three. Ramus summation applied for instance to the middle term yields

$$\frac{1}{2^{p+1}}\sum_{0\le k<2^{p+1}}\omega^{-k(n-2^p)}(1+\omega^k)^{2n}=\sum_{j\in\mathbf{Z}}\binom{2n}{n-2^p+j2^{p+1}},$$

so that translating the index in the sum and dealing accordingly with the other two sums:

$$R_{p,n}=\sum_{\substack{j\in\mathbb{Z}\\ q=2^p(2j+1)}}\left[\binom{2n}{n+1+q}-2\binom{2n}{n+q}+\binom{2n}{n-1-q}\right].$$

This expression is clearly equivalent to that of the theorem; expressions for $S_{p,n}$ and M_n result immediately from summation.

3. Asymptotic estimations

This section is devoted to a detailed asymptotic analysis of the quantity $A_n = M_n/B_n$ which represents the average number of registers needed to evaluate trees of size n. We start with the value of M_n provided by the last section:

$$M_n=\sum_{i>0} v_2(i)\left[\binom{2n}{n+i+1}-2\binom{2n}{n+i}+\binom{2n}{n+i-1}\right].$$

This formula involves binomial coefficients on one hand and the dyadic valuation v_2 on the other. The erratic behaviour of v_2 is smoothed by means of successive summations (Section 3.1). The Gaussian approximation of binomial coefficients is given in Section 3.2. Summation also leads to an expression involving the sum-of-digits function for which direct asymptotic information is available; actually an exact expression for this function has been given by Delange [3] whose results are reviewed in Section 3.3.

An asymptotic expansion of A_n is then obtained in Section 3.4, in which appears a periodic term; its Fourier series expansion is studied in Section 3.5.

3.1. Smoothing through summation by parts

At this point, it is convenient to introduce some operators from the calculus of finite differences (see [8]):

shift operator E: $(Ef)(k)=f(k-1)$,

difference operator Δ: $(\Delta f)(k)=f(k)-f(k-1)$,

sum operator Σ: $(\Sigma f)(k)=\sum_{0<i\le k} f(i)$.

In this notation, Theorem 1 can be expressed by

$$M_n=\Sigma v_2(i)\Delta^2\binom{2n}{n+i+1}.$$

The formula of summation by parts, also known as Abel's transformation, reads

$$\Sigma(\Delta f)g=fg-\Sigma(Ef)(\Delta g).$$

Applying it to M_n gives:

$$M_n = -\sum_{i>0} \sum_{0<j\leq i} v_2(j)\Delta^3\binom{2n}{n+i+2}.$$

In order to simplify the term $\sum_{0<j\leq i} v_2(j)$, consider the binary representations of two consecutive integers $j-1$ and j, for $j \geq 1$:

$$\mathrm{BIN}(j-1) = \mu\ 0\ 1\ 1\ \cdots\ 1 \quad \text{and}$$

$$\mathrm{BIN}(j) \quad = \mu\ 1\ 0\ 0\ \cdots\ 0.$$

with $\mu \in \{0, 1\}^*$. The number of trailing zeroes in $\mathrm{BIN}(j)$ is precisely $v_2(j)$. If we let $S_2(j)$ denote the sum of the digits of j in its binary representation, i.e. the number of ones in $\mathrm{BIN}(j)$, this observation shows $S_2(j-1) - S_2(j) = v_2(j) - 1$ thus: $\sum_{0<j\leq i} v_2(j) = i - S_2(i)$. (Remark that, since v_2 is multiplicative, i.e. $v_2(r) + v_2(s) = v_2(r \cdot s)$ this last expression is also $v_2(i!) = i - S_2(i)$, the exponent of 2 in the prime decomposition of i factorial).

Substituting in the expression for M_n yields

$$M_n = B_n + \sum_{i>0} S_2(i)\Delta^3\binom{2n}{n+i+2},$$

where

$$B_n = \binom{2n}{n} - \binom{2n}{n+1} = \frac{1}{n+1}\binom{2n}{n}.$$

Another summation by parts leads to

$$M_n = B_n - \sum_i \Big(\sum_{0<j<i} S_2(j)\Big)\, \Delta^4\binom{2n}{n+i+2}.$$

Thus, the problem of estimating M_n reduces, on the one hand to that of estimating

$$T(i) = \sum_{0<j<i} S_2(j)$$

and, on the other, to estimating finite differences of binomial coefficients.

3.2. *Asymptotic equivalent of binomial coefficients*

We use here an approximation by Gaussian distribution, well-known in probability theory (see [5]):

$$\binom{2n}{n+k} = \binom{2n}{n} \mathrm{e}^{-k^2/n}\left(1 + \mathrm{O}\left(\frac{(\log n)^*}{\sqrt{n}}\right)\right),$$

for $0 \leq k \leq \sqrt{n}\log n$. Here $\mathrm{O}(x^*)$ is to be interpreted as $\mathrm{O}(x^k)$ for some $k \in \mathbf{N}$, and $\log n$ denotes, from now on, the logarithm of n with base 2.

Let $H_j(t) = e^{t^2}(d^j/dt^j)(e^{-t^2})$ denote the jth Hermite polynomial; thus $H_0 = 1$, $H_1 = 2t$, $H_2 = 4t^2 - 2$, $H_3 = -8t^3 + 12t$, $H_4 = 16t^4 - 48t^2 + 12, \cdots$. An elementary computation shows that:

Proposition 6. *The ith finite difference of a binomial coefficient is estimated by*

$$\Delta^i \binom{2n}{n+k} = \binom{2n}{n} \frac{e^{-k^2/n}}{n^{i/2}} \left[H_i\left(\frac{k}{\sqrt{n}}\right) + O\left(\frac{(\log n)^*}{\sqrt{n}}\right)\right]$$

for $|k| \le \sqrt{n} \log n$.

In particular, one can check that:

$$\Delta^4 \binom{2n}{n+k} = \frac{1}{n^2}\binom{2n}{n} e^{-t^2}\left[H_4(t) + O\left(\frac{(\log n)^*}{\sqrt{n}}\right)\right]$$

for $t = k/\sqrt{n}$ and $|t| \le \log n$.

3.3. Number of ones in binary expansions of integers

In this section, we briefly recall a result by [3] which gives a simple form for the function $T(n)$ representing the number of ones in the binary expansions of all integers $0, 1, \cdots, n-1$.

Theorem 2 (Delange). *There exists a real function F such that for all* $n \ge 0$

$$T(n) = \tfrac{1}{2} n \log n + nF(\log n)$$

with F continuous and periodic of period 1.

Proof (Sketch). Take the binary representations of integers $0, 1, \ldots, n-1$. Each of these has length bounded by $1 + \lfloor \log n \rfloor$.

Let $t_r(n)$ be the total number of ones at position r in the representation of integers $0, 1, \ldots, n-1$.

$$T(n) = \sum_{0 \le j \le \lfloor \log n \rfloor} t_r(n)$$

Notice that for fixed r, $t_r(n) \sim n/2$ as n gets large since each binary position consists of asymptotically as many zeroes and ones. This observation leads to the asymptotic equivalent

$$T(n) \sim \tfrac{1}{2} n \log n.$$

As shown by Delange, it is possible to refine on this evaluation by observing some periodicity in $t_r(n) - n/2$. Let $g(x)$ be the triangular function defined by

$$g(x) = \begin{cases} -\dfrac{x}{2} & \text{for } 0 \le x < \tfrac{1}{2} \\[2ex] \dfrac{x-1}{2} & \text{for } \tfrac{1}{2} \le x < 1 \end{cases}$$

and extended by periodicity $g(x) = g(x+1)$ for all x; it is seen that

$$i_r(n) - n/2 = 2^{r+1} g\left(\frac{n}{2^{r+1}}\right).$$

Summation leads to an expression for $T(n)$:

$$T(n) = \frac{n}{2}(1+l) + 2^{1+l} \sum_{j \geq 0} 2^{-j} g\left(2^j \frac{n}{2^{1+l}}\right),$$

where $l = \lfloor \log n \rfloor$.

Defining $h(x) = \sum_{j \geq 0} 2^{-j} g(2^j x)$, we have

$$T(n) = \frac{n}{2}(1+l) + 2^{1+l} h\left(\frac{n}{2^{1+l}}\right).$$

Function h appears as a superposition of triangular functions, as suggested in Fig. 5.

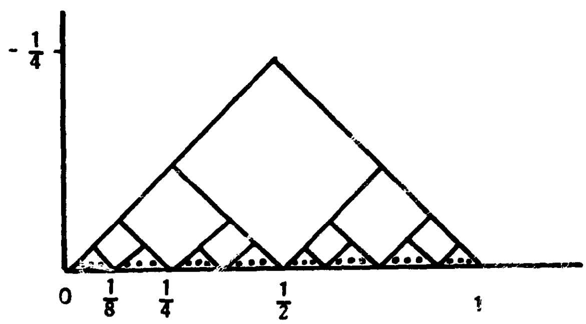

Fig. 5.

Now, the expression for F is transformed into $T(n) = \frac{1}{2} n \log n + nF(\log n)$ by taking $F(u) = \frac{1}{2}c(u) + 2^{c(u)} h(2^{-c(u)})$, where $c(u) = 1 + \lfloor u \rfloor - u$ is the complement of the fractional part of u.

Delange has shown that function F is nowhere differentiable and has determined its Fourier expansion, a result which we use in Section 3.5 (see [23] for a similar non differentiable continuous function).

3.4. The asymptotic estimation of M_n

Recall that

$$M_n = B_n - \sum_i T(i) \Delta^4 \binom{2n}{n+i+2}$$

in which $T(i) = \frac{1}{2} i \log i + iF(\log i)$. In Section 3.2, we showed that

$$\Delta^4 \binom{2n}{n+i+2} = \binom{2n}{n} \frac{e^{-i^2/n}}{n^2} \left[H_4\left(\frac{i}{\sqrt{n}}\right) + O\left(\frac{\log^* n}{\sqrt{n}}\right) \right].$$

Moreover, the binomial coefficient $\binom{2n}{n+k}$ is exponentially small compared to $\binom{2n}{n}$ when $k \geq \sqrt{n}\log n$, so that the summation can be restricted to those k which satisfy $k < \sqrt{n}\log n$:

$$M_n = B_n - \sum_{i<\sqrt{n}\log(n)} \binom{2n}{n} T(i) \frac{e^{-i^2/n}}{n^2}\left[H_4\left(\frac{i}{\sqrt{n}}\right) + O\left(\frac{\log^* n}{n}\right)\right].$$

Let us evaluate the corrective term. Since $T(i) = O(i \log i)$, each term of the sum is of order

$$O\left(\binom{2n}{n}\frac{i \log i}{n^2}\frac{\log^* n}{\sqrt{n}}\right).$$

There are $O(\sqrt{n}\log n)$ such terms and $i < \sqrt{n}\log n$. Hence the corrective term is of order

$$O\left(\sqrt{n}\log n \binom{2n}{n}\frac{n \log n \log^* n}{n^2 \sqrt{n}}\right) = O\left(\frac{1}{n}\binom{2n}{n}\frac{(\log n)^*}{\sqrt{n}}\right)$$
$$= B_n O\left(\frac{(\log n)^*}{\sqrt{n}}\right).$$

In the remaining sum, set $t = i/\sqrt{n}$, $t = 0, 1/\sqrt{n}, \ldots, \log n$.

Thus, calling $\sigma = \{0, 1/\sqrt{n}, \ldots, \log n\}$ and replacing $T(i)$ by its value:

$$M_n = B_n - \frac{1}{n^2}\binom{2n}{n}\sum_{t\in\sigma}\left[\frac{t\sqrt{n}}{2}\log t\sqrt{n} + t\sqrt{n}F(\log t + \tfrac{1}{2}\log n)\right] e^{-t^2}H_4(t),$$

hence

$$\frac{M_n}{B_n} = 1 - \frac{\log n}{\sqrt{n}}\sum_{t\in\sigma}\frac{t}{4}H_4(t)\,e^{-t^2} - \frac{1}{\sqrt{n}}\sum_{t\in\sigma}\frac{t}{2}(\log t)H_4(t)\,e^{-t^2}$$
$$-\frac{1}{\sqrt{n}}\sum_{t\in\sigma} tF(\log t + \tfrac{1}{2}\log n)H_4(t)\,e^{-t^2} + O\left(\frac{\log^* n}{\sqrt{n}}\right).$$

Each of these three sums is a Riemann sum on the interval $[0, \log n]$ for the corresponding function, and is known to approximate the integral of the function, with an error given by the following simple lemma.

Lemma. *Let f be a function on the interval $[a, b]$ and $a = x_0, \ldots, x_n = b$ a sub-division of the interval. If f possesses an integral over $[a, b]$, then*

$$\left|\int_a^b f(x)\,dx - \sum_{0<i\leq n}(x_i - x_{i-1})f(x_i)\right| \leq (b-a)\sup_i \operatorname{osc}(f; [x_{i-1}, x_i]),$$

where the oscillation of f over the interval I is defined by $\operatorname{osc}(f; I) = \sup f(I) - \inf f(I)$.

Proof. Decompose the integral into a sum:

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{0<i\leq n} \int_{x_{i-1}}^{x_i} f(x)\,\mathrm{d}x.$$

For $x_{i-1}\leq x\leq x_i$, $f(x)=f(x_i)+\varepsilon_i$ and

$$|\varepsilon_i|\leq \operatorname{osc}(f;[x_{i-1},x_i])\leq \sup_i \operatorname{osc}(f;[x_{i-1},x_i])=\omega.$$

It follows that

$$\int_a^b f(x)\,\mathrm{d}x = \sum_{0<i\leq n}(x_i-x_{i-1})f(x_i)+\sum_{0<i\leq n}\int_{x_{i-1}}^{x_i}\varepsilon_i\,\mathrm{d}x;$$

thus the corrective term satisfies

$$\left|\sum_{0<i\leq n}\int_{x_{i-1}}^{x_i}\varepsilon_i\,\mathrm{d}x\right|\leq \sum_{0<i\leq n}\left|\int_{x_{i-1}}^{x_i}\varepsilon_i\,\mathrm{d}x\right|$$

$$\leq \sum_{0<i\leq n}\int_{x_{i-1}}^{x_i}|\varepsilon_i|\,\mathrm{d}x\leq \sum_{0<i\leq n}(x_i-x_{i-1})=(b-a)\omega.$$

In particular, when f is differentiable with derivative bounded by M, then $\operatorname{osc}(f;[x_{i-1},x_i])\leq(x_i-x_{i-1})M$. In the case of our first two sums, the error is thus bounded by $O((\log n)/\sqrt{n}^*)$. Unfortunately, F is not differentiable and requires a special treatment.

Recall that $F(u)=\frac{1}{2}c(u)+2^{c(u)}h(2^{-c(u)})$ in which

$$h(x)=\sum_{j\geq 0}2^{-j}g(2^jx).$$

The functions $\frac{1}{2}c(u)$, $2^{c(u)}$ and $x=2^{-c(u)}$ are differentiable on the interval $[0, 1[$ with bounded derivative. Their oscillations on an interval of length $1/\sqrt{n}$ are thus of order $O(1/\sqrt{n})$. If h is also of oscillation $O((\log n/\sqrt{n})$, then F is of oscillation $O((\log n)/\sqrt{n})$. Since

$$h(x)=\sum_{0\leq j<\log n}2^{-j}g(2^jx)+\sum_{j\geq\log n}2^{-j}g(2^jx)$$

$$h(x)=\sum_{0\leq j<\log n}2^{-j}g(2^jx)+2^{-\lfloor\log n\rfloor}h(2^{\lfloor\log n\rfloor}x),$$

function h appears as the sum of two functions: the second one is of order $O(1/n)$ and certainly has oscillation of order $O(1/n)$; the first one is in turn the sum of $\log n$ functions each of which having oscillation $O(1/\sqrt{n})$, and thus has oscillation $O((\log n)/\sqrt{n})$, so that eventually,

$$\frac{M_n}{B_n}=1-(\log n)I_1-I_2-I_3+O\left(\frac{(\log n)^*}{\sqrt{n}}\right)$$

in which

$$I_1 = \int_0^{\log n} \frac{t}{4} e^{-t^2} H_4(t)\, dt$$

$$I_2 = \int_0^{\log n} \frac{t \log t}{2} e^{-t^2} H_4(t)\, dt$$

$$I_3 = \int_0^{\log n} tF(\log t + \tfrac{1}{2}\log n) e^{-t^2} H_4(t)\, dt.$$

Each of the integrand is of order $O(e^{-t^2} H_6(t))$, so that the integrals can be pushed to infinity with error exponentially small.

$$\frac{M_n}{B_n} = 1 - (\log n) J_1 - J_2 - J_3 + O\left(\frac{(\log^* n)}{\sqrt{n}}\right),$$

where J_1, J_2, J_3 are computed in turn:

$$J_1 = \int_0^{\infty} \frac{t}{4} e^{-t^2} H_4(t)\, dt = -\tfrac{1}{2},$$

$$J_2 = \int_0^{\infty} \frac{t \log t}{2} e^{-t^2} H_4(t)\, dt = \frac{\gamma}{2 \ln 2},$$

$$J_3 = \int_0^{\infty} tF(\log t + \tfrac{1}{2}\log n)\, e^{-t^2} H_4(t)\, dt = K(\log_4 n).$$

We have just proved

Theorem 3. The average number of registers $A_n = M_n / B_n$ satisfies

$$A_n = \log_4 n + D(\log_4 n) + o(1),$$

in which

$$D(u) = 1 - \frac{\gamma}{2 \ln 2} - \int_0^{\infty} tH_4(t) F(\log t + u)\, e^{-t^2}\, dt,$$

is a continuous function with period one, and F is the function in Theorem 2.

3.5. The periodicity in the asymptotic expansion

Results from the last section have revealed an oscillating term in the asymptotic expansion of the average number of register A_n. In this section, we show this periodic term has a Fourier expansion and we compute its coefficients.

This computation is based on the possibility of expanding F as shown in [3]: function F is the sum of terms, each of which is expressed in terms of the triangular function $g(x)$ and has a Fourier expansion. So that:

Lemma. *The function F has a Fourier expansion $F(u)=\Sigma f_k\, e^{i2k\pi u}$, where the coefficients are given by*

$$f_0=\tfrac{1}{2}\log \pi-\frac{\gamma}{2\ln 2}-\frac{1}{4}$$

$$f_k=-\frac{1}{\ln 2}\chi_k^{-1}(1+\chi_k)^{-1}\zeta(\chi_k)\quad \text{with } \chi_k=\frac{2k\pi i}{\ln 2}.$$

We prove:

Corollary. *The average number of registers to evaluate expressions of size n is given by $A_n=\log_4 n+D(\log_4 n)+o(1)$ where D has the Fourier expansion $D(u)=\Sigma d_k\, e^{i2k\pi u}$, and the coefficients d_k are given by*

$$d_0=\frac{1}{2}-\frac{\gamma+2}{2\ln 2}+\log \pi\simeq 0.292\cdots,$$

$$d_k=\frac{1}{\ln 2}(\chi_k-1)\Gamma\left(\frac{\chi_k}{2}\right)\zeta(\chi_k)\quad \text{with } \chi_k=\frac{2k\pi i}{\ln 2}.$$

Proof. We start from

$$D(u)=1-\frac{\gamma}{2\ln 2}-\int_0^\infty tH_4(t)F(u+\log t)\, e^{-t^2}\, dt.$$

The problem thus amounts to computing the Fourier coefficients of the integral. This integral is a variant of a convolution product of $tH_4(t)\, e^{-t^2}$, whose expansion involves the gamma function $\Gamma(t)$, and of $F(v)$, whose expansion involves the zeta function $\zeta(t)$.

Let

$$a_k=\int_0^1 e^{-2ik\pi u}\int_0^\infty F(\log t+u)H_4(t)t\, e^{-t^2}\, dt\, du.$$

The double integral exists and we are justified in interchanging the order of summations:

$$a_k=\int_0^\infty H_4(t)t\, e^{-t^2}\int_0^1 e^{-2ik\pi u}F(\log t+u)\, du\, dt.$$

We let $v=\log t+u$ in the innermost integral and using the periodicity of F we obtain:

$$a_k=\int_0^\infty H_4(t)t\, e^{-t^2}\, e^{2ik\pi\log t}\int_0^1 e^{-2i\pi kv}F(v)\, dv\, dt.$$

As expected, the computation of a_k splits into two independent parts:

$$f_k = \int_0^1 e^{-2ik\pi v} F(v)\, dv \text{ is provided by Delange's expansion of } F;$$

$$c_k = \int_0^\infty H_4(t) t\, e^{-t^2} e^{2ik\pi \log t}\, dt \text{ is computed by setting } t^2 = u.$$

Hence

$$c_k = \frac{1}{2}\int_0^\infty (16u^2 - 48u + 12)\, e^{-u} u^{\frac{ik\pi}{\ln 2}}\, du$$

$$= \frac{1}{2}\left[16\Gamma\left(3+\frac{ik\pi}{\ln 2}\right) - 48\Gamma\left(2+\frac{ik\pi}{\ln 2}\right) + 12\Gamma\left(1+\frac{ik\pi}{\ln 2}\right)\right]$$

For $k = 0$, $c_k = 2$; For $k \neq 0$, the basic functional property of the gamma function shows that

$$c_k = \chi_k(\chi_k^2 - 1)\Gamma\left(\frac{\chi_k}{2}\right) \quad \text{with } \chi_k = \frac{2k\pi i}{\ln 2}.$$

Putting things together, we get:

$$d_0 = 1 - \frac{\gamma}{2\ln 2} - c_0 \cdot f_0$$

$$= \frac{1}{2} - \frac{\gamma}{2\ln 2} - \frac{1}{\ln 2} + \log \pi$$

$$d_k = c_k f_k = \frac{1}{\ln 2}(\chi_k - 1)\Gamma\left(\frac{\chi_k}{2}\right)\zeta(\chi_k) \quad \text{for } k \neq 0.$$

Finally, the Fourier series of D is absolutely convergent since $\zeta(i\chi) = O(|\chi|^{1/2+\varepsilon})$ and $\Gamma(i\chi) = O(e^{-\pi|\chi|})$ when $\chi \to \infty$ (cf [24]).

4. Conclusions

We propose here a few observations on the significance of results obtained in the last sections.

4.1. *The average number of registers*

Considering only the first terms of asymptotic expansions, we see that the evaluation of expressions of size n requires

on the average $\sim \log_4 n$ registers,

on the maximum $\sim \log_2 n$ registers.

Thus on the average, one needs significantly less (one half) registers than in the worst case; this suggests possible economies in register allocation strategies (see below for distribution results).

We can first refine on this observation by taking a closer look at the asymptotic result given by Theorem 3, especially for small values of n. Fig. 6 represents the values of $A_n - \log_4 n$ for $2 \leq n \leq 300$, and shows excellent agreement with Theorem 3: the corrected term quickly conforms to a periodic term centered around $d_0 \simeq 0.292 \cdots$. The periodic term is seen to have amplitude bounded by 0.05 which is confirmed by the Fourier expansion given in the Corollary of Section 3.5.

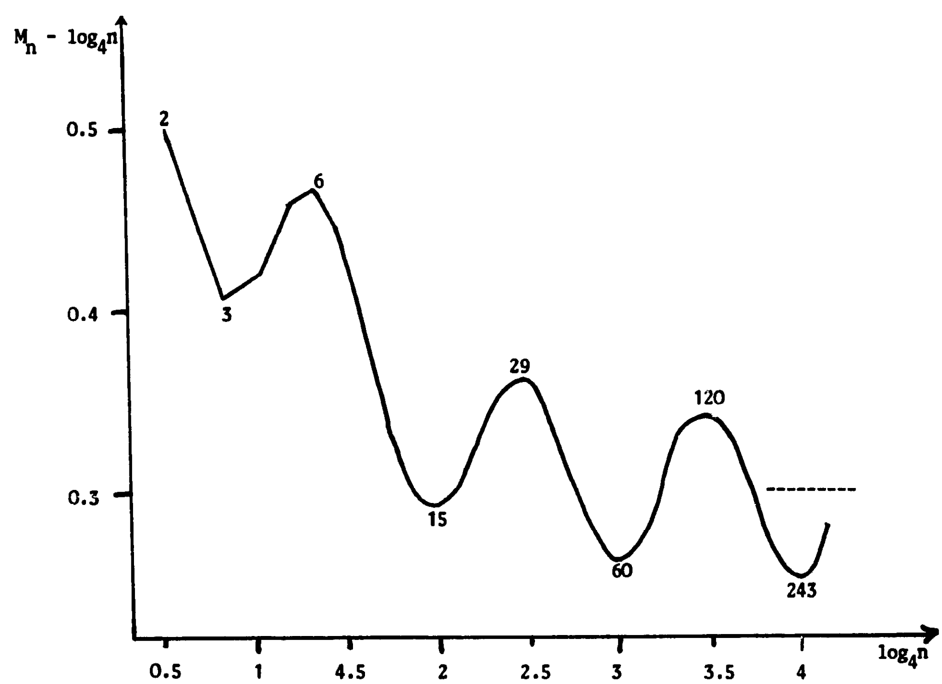

Fig. 6. Values of $A_n - \log_4 n$.

4.2. *The distribution of the register functions*

Results from Section 2 show that for expressions of a given size n the register function has a strong peak around $\log_4 n$. For instance when $n = 40$, some expressions require as much as 4 registers; however 96% can be evaluated with 2 registers and more than 99.9% can be evaluated with only three. Similar distribution results are summarized in Table 3 where for each size n, and for each register number p we indicate the proportion of expressions of size n requiring p registers. As already indicated, this table shows that register allocation algorithms can use a few registers less than the maximum without detectible loss in efficiency; in any case it makes it possible to precisely evaluate the trade-offs between space and time for limited memory allocation algorithms.

Table 3. The distribution of expressions by number of registers $\frac{R_{p,n}}{B_n}$

n \ p	0	1	2	3	4	5
3	80%	20%				
5	38%	62%				
10	3%	92%	5%			
20	0%	52%	48%	0%		
40	0%	6%	90%	4%	0%	
100	0%	0%	35%	65%	0%	0%

4.3. Exploration of trees

As shown in [1], the average stack height to explore trees of size n is $\sqrt{\pi n}+O(1)$; we have shown that for explorations where distinction between left and right is immaterial the stack height is reduced to an average of $\log_4 n + O(1)$.

4.4. Analysis of algorithms

Several analyses of algorithms (exploration of trees, register allocation, odd–even merge $\cdots$) have an average case behaviour described by convolution products of binomial coefficients and some arithmetic functions. Some of the asymptotic evaluations have been performed by use of complex integral transforms: indeed [10] has independently derived Theorem 3 in this way. We show here the possibility of a more elementary derivation where a periodicity in the asymptotic expansion is seen to originate in a similar periodicity in the arithmetical function involved in the analysis. In this respect, striking similarities exist between the analyses of odd-even merge in another independent work of [18] and of register allocation. This last analysis can also be subjected to the same treatment as shown in [7]; the periodicity there is intrinsic in the Gray code representation of numbers.

Note. R. Kemp has independently established several of these results using different methods.

Acknowledgements

We thank H. Delange for explicitly showing us how to derive the Fourier expansion of Section 3.5. N.G. de Bruijn by an early result on a related problem stimulated our interest. M. Fiolet and J.L. Rémy both communicated interesting observations on the problem. We thank R. Sethi, E. Reingold and P. Downey for communicating bibliographical references.

References

[1] N.G. de Bruijn, D.E. Knuth and S.O. Rice, The average height of planted plane trees, *Graph Theory and Computing*, R.C. Read, ed., (Academic Press 1972) 15–22.
[2] M.F. Dacey, Summary of magnitude properties of topologically distinct channel networks and network patterns, in: Daniel F. Merriam, ed., *Random Processes in Geology* (Springer-Verlag, Berlin 1976) 16–38.
[3] H. Delange, Sur la fonction sommatoire de la fonction somme des chiffres, *Enseignement Math.* **21**(1) (1975).
[4] A.P. Ershov, On programming of arithmetic operations, *CACM* **1** (8) (1958) 3–6.
[5] W. Feller, *An introduction to probability theory and its applications*, Vol. 1 (Wiley, New York 1957).
[6] P. Flajolet and L. Ramshaw, A note on Gray-code and odd-even merge (1977), to appear.
[7] P. Flajolet, J. Françon, G. Viennot and J. Vuillemin, Arbres et Permutations: Combinatoire et Algorithmique, (1978), in preparation.
[8] R.E. Jordan, *Calculus of Finite Differences*, (Chelsea, New-York 1965).
[10] R. Kemp, The average number of registers to evaluate a binary tree optimally, Saarbrücken, University Report (1977).
[11] D.E. Knuth, *The Art of Computer Programming*, Vol. 1, (Addison-Wesley 1968) 315 sq.
[12] D.E. Knuth, Structured programming with go to statements, Comput. Surveys, **6** (4) (1974) 216–302.
[13] G. Kreweras, Sur les éventails de segments, Cahiers B.U.R.O. **15** (1970) 1–41.
[14] E. Lucas, *Theorie des Nombres* (Paris, 1891).
[15] T.A. McMahon, The mechanical design of trees, *Sci. Amer.* **233** (1) (1975) 92–102.
[16] I. Nakata, On compiling algorithms for arithmetic expressions, *CACM* **10** (8) (1967) 492–494.
[17] J. Riordan, Combinatorial identities (J. Wiley, New York 1968).
[18] R. Sedgewick, Data movement in odd-even merging, in: Proc. of the Conference on Theoretical Computer Science, Waterloo, 1977.
[19] R. Sethi and J.D. Ullman, The generation of optimal code for arithmetic expressions, *JACM* **17** 4 (1970) 715–728.
[20] R.L. Shreve, Statistical Law of stream numbers, *Geology*, **74** (1966) 17–37.
[21] P.S. Stevens, *Patterns in Nature* (Atlantic Monthly Press, Little Brown and Co., 1974) 108–116.
[22] A.N. Strahler, Hypsometric (area-altitude) analysis of erosional topography, *Bulletin Geological Society of America*, **63** (1952) 1117–1142.
[23] B.L. van der Waerden, Ein einfaches Beispiel einer nicht-differenzierbaren stetigen Funktion, *Math. Z*, **32** (1930) 474–475.
[24] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, 1902).
[25] M.J. Woldenberg, A structural taxonomy of spatial hierarchies, *Colston Papers*, **22**, pp. 147–175, London: Butterworths Scientific Publishers, 1970.